

Network Working Group
Request for Comments: 2382
Category: Informational

E. Crawley, Editor
Argon Networks
L. Berger
Fore Systems
S. Berson
ISI
F. Baker
Cisco Systems
M. Borden
Bay Networks
J. Krawczyk
ArrowPoint Communications
August 1998

A Framework for Integrated Services and RSVP over ATM

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (1998). All Rights Reserved.

Abstract

This document outlines the issues and framework related to providing IP Integrated Services with RSVP over ATM. It provides an overall approach to the problem(s) and related issues. These issues and problems are to be addressed in further documents from the ISATM subgroup of the ISSLL working group.

1. Introduction

The Internet currently has one class of service normally referred to as "best effort." This service is typified by first-come, first-serve scheduling at each hop in the network. Best effort service has worked well for electronic mail, World Wide Web (WWW) access, file transfer (e.g. ftp), etc. For real-time traffic such as voice and video, the current Internet has performed well only across unloaded portions of the network. In order to provide quality real-time traffic, new classes of service and a QoS signalling protocol are

being introduced in the Internet [1,6,7], while retaining the existing best effort service. The QoS signalling protocol is RSVP [1], the Resource ReSerVation Protocol and the service models

One of the important features of ATM technology is the ability to request a point-to-point Virtual Circuit (VC) with a specified Quality of Service (QoS). An additional feature of ATM technology is the ability to request point-to-multipoint VCs with a specified QoS. Point-to-multipoint VCs allows leaf nodes to be added and removed from the VC dynamically and so provides a mechanism for supporting IP multicast. It is only natural that RSVP and the Internet Integrated Services (IIS) model would like to utilize the QoS properties of any underlying link layer including ATM, and this memo concentrates on ATM.

Classical IP over ATM [10] has solved part of this problem, supporting IP unicast best effort traffic over ATM. Classical IP over ATM is based on a Logical IP Subnetwork (LIS), which is a separately administered IP subnetwork. Hosts within an LIS communicate using the ATM network, while hosts from different subnets communicate only by going through an IP router (even though it may be possible to open a direct VC between the two hosts over the ATM network). Classical IP over ATM provides an Address Resolution Protocol (ATMARP) for ATM edge devices to resolve IP addresses to native ATM addresses. For any pair of IP/ATM edge devices (i.e. hosts or routers), a single VC is created on demand and shared for all traffic between the two devices. A second part of the RSVP and IIS over ATM problem, IP multicast, is being solved with MARS [5], the Multicast Address Resolution Server.

MARS compliments ATMARP by allowing an IP address to resolve into a list of native ATM addresses, rather than just a single address.

The ATM Forum's LAN Emulation (LANE) [17, 20] and Multiprotocol Over ATM (MPOA) [18] also address the support of IP best effort traffic over ATM through similar means.

A key remaining issue for IP in an ATM environment is the integration of RSVP signalling and ATM signalling in support of the Internet Integrated Services (IIS) model. There are two main areas involved in supporting the IIS model, QoS translation and VC management. QoS translation concerns mapping a QoS from the IIS model to a proper ATM QoS, while VC management concentrates on how many VCs are needed and which traffic flows are routed over which VCs.

1.1 Structure and Related Documents

This document provides a guide to the issues for IIS over ATM. It is intended to frame the problems that are to be addressed in further documents. In this document, the modes and models for RSVP operation over ATM will be discussed followed by a discussion of management of ATM VCs for RSVP data and control. Lastly, the topic of encapsulations will be discussed in relation to the models presented.

This document is part of a group of documents from the ISATM subgroup of the ISSLL working group related to the operation of IntServ and RSVP over ATM. [14] discusses the mapping of the IntServ models for Controlled Load and Guaranteed Service to ATM. [15 and 16] discuss detailed implementation requirements and guidelines for RSVP over ATM, respectively. While these documents may not address all the issues raised in this document, they should provide enough information for development of solutions for IntServ and RSVP over ATM.

1.2 Terms

Several term used in this document are used in many contexts, often with different meaning. These terms are used in this document with the following meaning:

- Sender is used in this document to mean the ingress point to the ATM network or "cloud".
- Receiver is used in this document to refer to the egress point from the ATM network or "cloud".
- Reservation is used in this document to refer to an RSVP initiated request for resources. RSVP initiates requests for resources based on RESV message processing. RESV messages that simply refresh state do not trigger resource requests. Resource requests may be made based on RSVP sessions and RSVP reservation styles. RSVP styles dictate whether the reserved resources are used by one sender or shared by multiple senders. See [1] for details of each. Each new request is referred to in this document as an RSVP reservation, or simply reservation.
- Flow is used to refer to the data traffic associated with a particular reservation. The specific meaning of flow is RSVP style dependent. For shared style reservations, there is one flow per session. For distinct style reservations, there is one flow per sender (per session).

2. Issues Regarding the Operation of RSVP and IntServ over ATM

The issues related to RSVP and IntServ over ATM fall into several general classes:

- How to make RSVP run over ATM now and in the future
- When to set up a virtual circuit (VC) for a specific Quality of Service (QoS) related to RSVP
- How to map the IntServ models to ATM QoS models
- How to know that an ATM network is providing the QoS necessary for a flow
- How to handle the many-to-many connectionless features of IP multicast and RSVP in the one-to-many connection-oriented world of ATM

2.1 Modes/Models for RSVP and IntServ over ATM

[3] Discusses several different models for running IP over ATM networks. [17, 18, and 20] also provide models for IP in ATM environments. Any one of these models would work as long as the RSVP control packets (IP protocol 46) and data packets can follow the same IP path through the network. It is important that the RSVP PATH messages follow the same IP path as the data such that appropriate PATH state may be installed in the routers along the path. For an ATM subnetwork, this means the ingress and egress points must be the same in both directions for the RSVP control and data messages. Note that the RSVP protocol does not require symmetric routing. The PATH state installed by RSVP allows the RESV messages to "retrace" the hops that the PATH message crossed. Within each of the models for IP over ATM, there are decisions about using different types of data distribution in ATM as well as different connection initiation. The following sections look at some of the different ways QoS connections can be set up for RSVP.

2.1.1 UNI 3.x and 4.0

In the User Network Interface (UNI) 3.0 and 3.1 specifications [8,9] and 4.0 specification, both permanent and switched virtual circuits (PVC and SVC) may be established with a specified service category (CBR, VBR, and UBR for UNI 3.x and VBR-rt and ABR for 4.0) and specific traffic descriptors in point-to-point and point-to-multipoint configurations. Additional QoS parameters are not available in UNI 3.x and those that are available are vendor-specific. Consequently, the level of QoS control available in standard UNI 3.x networks is somewhat limited. However, using these building blocks, it is possible to use RSVP and the IntServ models. ATM 4.0 with the Traffic Management (TM) 4.0 specification [21] allows much greater control of QoS. [14] provides the details of mapping the IntServ models to UNI 3.x and 4.0 service categories and traffic parameters.

2.1.1.1 Permanent Virtual Circuits (PVCs)

PVCs emulate dedicated point-to-point lines in a network, so the operation of RSVP can be identical to the operation over any point-to-point network. The QoS of the PVC must be consistent and equivalent to the type of traffic and service model used. The devices on either end of the PVC have to provide traffic control services in order to multiplex multiple flows over the same PVC. With PVCs, there is no issue of when or how long it takes to set up VCs, since they are made in advance but the resources of the PVC are limited to what has been pre-allocated. PVCs that are not fully utilized can tie up ATM network resources that could be used for SVCs.

An additional issue for using PVCs is one of network engineering. Frequently, multiple PVCs are set up such that if all the PVCs were running at full capacity, the link would be over-subscribed. This frequently used "statistical multiplexing gain" makes providing IIS over PVCs very difficult and unreliable. Any application of IIS over PVCs has to be assured that the PVCs are able to receive all the requested QoS.

2.1.1.2 Switched Virtual Circuits (SVCs)

SVCs allow paths in the ATM network to be set up "on demand". This allows flexibility in the use of RSVP over ATM along with some complexity. Parallel VCs can be set up to allow best-effort and better service class paths through the network, as shown in Figure 1. The cost and time to set up SVCs can impact their use. For example, it may be better to initially route QoS traffic over existing VCs until a SVC with the desired QoS can be set up for the flow. Scaling issues can come into play if a single RSVP flow is used per VC, as will be discussed in Section 4.3.1.1. The number of VCs in any ATM device may also be limited so the number of RSVP flows that can be supported by a device can be strictly limited to the number of VCs available, if we assume one flow per VC. Section 4 discusses the topic of VC management for RSVP in greater detail.

2.1.2 Hop-by-Hop vs. Short Cut

If the ATM "cloud" is made up a number of logical IP subnets (LISs), then it is possible to use "short cuts" from a node on one LIS directly to a node on another LIS, avoiding router hops between the LISs. NHRP [4], is one mechanism for determining the ATM address of the egress point on the ATM network given a destination IP address. It is a topic for further study to determine if significant benefit is achieved from short cut routes vs. the extra state required.

2.1.3 Future Models

ATM is constantly evolving. If we assume that RSVP and IntServ applications are going to be wide-spread, it makes sense to consider changes to ATM that would improve the operation of RSVP and IntServ over ATM. Similarly, the RSVP protocol and IntServ models will continue to evolve and changes that affect them should also be considered. The following are a few ideas that have been discussed that would make the integration of the IntServ models and RSVP easier or more complete. They are presented here to encourage continued development and discussion of ideas that can help aid in the integration of RSVP, IntServ, and ATM.

2.1.3.1 Heterogeneous Point-to-MultiPoint

The IntServ models and RSVP support the idea of "heterogeneous receivers"; e.g., not all receivers of a particular multicast flow are required to ask for the same QoS from the network, as shown in Figure 2.

The most important scenario that can utilize this feature occurs when some receivers in an RSVP session ask for a specific QoS while others receive the flow with a best-effort service. In some cases where there are multiple senders on a shared-reservation flow (e.g., an audio conference), an individual receiver only needs to reserve enough resources to receive one sender at a time. However, other receivers may elect to reserve more resources, perhaps to allow for some amount of "over-speaking" or in order to record the conference (post processing during playback can separate the senders by their source addresses).

In order to prevent denial-of-service attacks via reservations, the service models do not allow the service elements to simply drop non-conforming packets. For example, Controlled Load service model [7] assigns non-conformant packets to best-effort status (which may result in packet drops if there is congestion).

Emulating these behaviors over an ATM network is problematic and needs to be studied. If a single maximum QoS is used over a point-to-multipoint VC, resources could be wasted if cells are sent over certain links where the reassembled packets will eventually be dropped. In addition, the "maximum QoS" may actually cause a degradation in service to the best-effort branches.

The term "variegated VC" has been coined to describe a point-to-multipoint VC that allows a different QoS on each branch. This approach seems to match the spirit of the Integrated Service and RSVP models, but some thought has to be put into the cell drop strategy when traversing from a "bigger" branch to a "smaller" one. The "best-effort for non-conforming packets" behavior must also be retained. Early Packet Discard (EPD) schemes must be used so that all the cells for a given packet can be discarded at the same time rather than discarding only a few cells from several packets making all the packets useless to the receivers.

2.1.3.2 Lightweight Signalling

Q.2931 signalling is very complete and carries with it a significant burden for signalling in all possible public and private connections. It might be worth investigating a lighter weight signalling mechanism for faster connection setup in private networks.

2.1.3.3 QoS Renegotiation

Another change that would help RSVP over ATM is the ability to request a different QoS for an active VC. This would eliminate the need to setup and tear down VCs as the QoS changed. RSVP allows receivers to change their reservations and senders to change their traffic descriptors dynamically. This, along with the merging of reservations, can create a situation where the QoS needs of a VC can change. Allowing changes to the QoS of an existing VC would allow these features to work without creating a new VC. In the ITU-T ATM specifications [24,25], some cell rates can be renegotiated or changed. Specifically, the Peak Cell Rate (PCR) of an existing VC can be changed and, in some cases, QoS parameters may be renegotiated during the call setup phase. It is unclear if this is sufficient for the QoS renegotiation needs of the IntServ models.

2.1.3.4 Group Addressing

The model of one-to-many communications provided by point-to-multipoint VCs does not really match the many-to-many communications provided by IP multicasting. A scaleable mapping from IP multicast addresses to an ATM "group address" can address this problem.

2.1.3.5 Label Switching

The MultiProtocol Label Switching (MPLS) working group is discussing methods for optimizing the use of ATM and other switched networks for IP by encapsulating the data with a header that is used by the interior switches to achieve faster forwarding lookups. [22] discusses a framework for this work. It is unclear how this work will affect IntServ and RSVP over label switched networks but there may be some interactions.

2.1.4 QoS Routing

RSVP is explicitly not a routing protocol. However, since it conveys QoS information, it may prove to be a valuable input to a routing protocol that can make path determinations based on QoS and network load information. In other words, instead of asking for just the IP next hop for a given destination address, it might be worthwhile for RSVP to provide information on the QoS needs of the flow if routing has the ability to use this information in order to determine a route. Other forms of QoS routing have existed in the past such as using the IP TOS and Precedence bits to select a path through the network. Some have discussed using these same bits to select one of a set of parallel ATM VCs as a form of QoS routing. ATM routing has also considered the problem of QoS routing through the Private Network-to-Network Interface (PNNI) [26] routing protocol for routing ATM VCs on a path that can support their needs. The work in this area is just starting and there are numerous issues to consider. [23], as part of the work of the QoSR working group frame the issues for QoS Routing in the Internet.

2.2 Reliance on Unicast and Multicast Routing

RSVP was designed to support both unicast and IP multicast applications. This means that RSVP needs to work closely with multicast and unicast routing. Unicast routing over ATM has been addressed [10] and [11]. MARS [5] provides multicast address resolution for IP over ATM networks, an important part of the solution for multicast but still relies on multicast routing protocols to connect multicast senders and receivers on different subnets.

2.3 Aggregation of Flows

Some of the scaling issues noted in previous sections can be addressed by aggregating several RSVP flows over a single VC if the destinations of the VC match for all the flows being aggregated. However, this causes considerable complexity in the management of VCs and in the scheduling of packets within each VC at the root point of

the VC. Note that the rescheduling of flows within a VC is not possible in the switches in the core of the ATM network. Virtual Paths (VPs) can be used for aggregating multiple VCs. This topic is discussed in greater detail as it applies to multicast data distribution in section 4.2.3.4

2.4 Mapping QoS Parameters

The mapping of QoS parameters from the IntServ models to the ATM service classes is an important issue in making RSVP and IntServ work over ATM. [14] addresses these issues very completely for the Controlled Load and Guaranteed Service models. An additional issue is that while some guidelines can be developed for mapping the parameters of a given service model to the traffic descriptors of an ATM traffic class, implementation variables, policy, and cost factors can make strict mapping problematic. So, a set of workable mappings that can be applied to different network requirements and scenarios is needed as long as the mappings can satisfy the needs of the service model(s).

2.5 Directly Connected ATM Hosts

It is obvious that the needs of hosts that are directly connected to ATM networks must be considered for RSVP and IntServ over ATM. Functionality for RSVP over ATM must not assume that an ATM host has all the functionality of a router, but such things as MARS and NHRP clients would be worthwhile features. A host must manage VCs just like any other ATM sender or receiver as described later in section 4.

2.6 Accounting and Policy Issues

Since RSVP and IntServ create classes of preferential service, some form of administrative control and/or cost allocation is needed to control access. There are certain types of policies specific to ATM and IP over ATM that need to be studied to determine how they interoperate with the IP and IntServ policies being developed. Typical IP policies would be that only certain users are allowed to make reservations. This policy would translate well to IP over ATM due to the similarity to the mechanisms used for Call Admission Control (CAC).

There may be a need for policies specific to IP over ATM. For example, since signalling costs in ATM are high relative to IP, an IP over ATM specific policy might restrict the ability to change the prevailing QoS in a VC. If VCs are relatively scarce, there also might be specific accounting costs in creating a new VC. The work so far has been preliminary, and much work remains to be done. The

policy mechanisms outlined in [12] and [13] provide the basic mechanisms for implementing policies for RSVP and IntServ over any media, not just ATM.

3. Framework for IntServ and RSVP over ATM

Now that we have defined some of the issues for IntServ and RSVP over ATM, we can formulate a framework for solutions. The problem breaks down to two very distinct areas; the mapping of IntServ models to ATM service categories and QoS parameters and the operation of RSVP over ATM.

Mapping IntServ models to ATM service categories and QoS parameters is a matter of determining which categories can support the goals of the service models and matching up the parameters and variables between the IntServ description and the ATM description(s). Since ATM has such a wide variety of service categories and parameters, more than one ATM service category should be able to support each of the two IntServ models. This will provide a good bit of flexibility in configuration and deployment. [14] examines this topic completely.

The operation of RSVP over ATM requires careful management of VCs in order to match the dynamics of the RSVP protocol. VCs need to be managed for both the RSVP QoS data and the RSVP signalling messages. The remainder of this document will discuss several approaches to managing VCs for RSVP and [15] and [16] discuss their application for implementations in term of interoperability requirement and implementation guidelines.

4. RSVP VC Management

This section provides more detail on the issues related to the management of SVCs for RSVP and IntServ.

4.1 VC Initiation

As discussed in section 2.1.1.2, there is an apparent mismatch between RSVP and ATM. Specifically, RSVP control is receiver oriented and ATM control is sender oriented. This initially may seem like a major issue, but really is not. While RSVP reservation (RESV) requests are generated at the receiver, actual allocation of resources takes place at the subnet sender. For data flows, this means that subnet senders will establish all QoS VCs and the subnet receiver must be able to accept incoming QoS VCs, as illustrated in Figure 1. These restrictions are consistent with RSVP version 1 processing rules and allow senders to use different flow to VC mappings and even different QoS renegotiation techniques without

interoperability problems.

The use of the reverse path provided by point-to-point VCs by receivers is for further study. There are two related issues. The first is that use of the reverse path requires the VC initiator to set appropriate reverse path QoS parameters. The second issue is that reverse paths are not available with point-to-multipoint VCs, so reverse paths could only be used to support unicast RSVP reservations.

4.2 Data VC Management

Any RSVP over ATM implementation must map RSVP and RSVP associated data flows to ATM Virtual Circuits (VCs). LAN Emulation [17], Classical IP [10] and, more recently, NHRP [4] discuss mapping IP traffic onto ATM SVCs, but they only cover a single QoS class, i.e., best effort traffic. When QoS is introduced, VC mapping must be revisited. For RSVP controlled QoS flows, one issue is VCs to use for QoS data flows.

In the Classic IP over ATM and current NHRP models, a single point-to-point VC is used for all traffic between two ATM attached hosts (routers and end-stations). It is likely that such a single VC will not be adequate or optimal when supporting data flows with multiple .bp QoS types. RSVP's basic purpose is to install support for flows with multiple QoS types, so it is essential for any RSVP over ATM solution to address VC usage for QoS data flows, as shown in Figure 1.

RSVP reservation styles must also be taken into account in any VC usage strategy.

This section describes issues and methods for management of VCs associated with QoS data flows. When establishing and maintaining VCs, the subnet sender will need to deal with several complicating factors including multiple QoS reservations, requests for QoS changes, ATM short-cuts, and several multicast specific issues. The multicast specific issues result from the nature of ATM connections. The key multicast related issues are heterogeneity, data distribution, receiver transitions, and end-point identification.

4.2.1 Reservation to VC Mapping

There are various approaches available for mapping reservations on to VCs. A distinguishing attribute of all approaches is how reservations are combined on to individual VCs. When mapping reservations on to VCs, individual VCs can be used to support a single reservation, or reservation can be combined with others on to

"aggregate" VCs. In the first case, each reservation will be supported by one or more VCs. Multicast reservation requests may translate into the setup of multiple VCs as is described in more detail in section 4.2.2. Unicast reservation requests will always translate into the setup of a single QoS VC. In both cases, each VC will only carry data associated with a single reservation. The greatest benefit of this approach is ease of implementation, but it comes at the cost of increased (VC) setup time and the consumption of greater number of VC and associated resources.

When multiple reservations are combined onto a single VC, it is referred to as the "aggregation" model. With this model, large VCs could be set up between IP routers and hosts in an ATM network. These VCs could be managed much like IP Integrated Service (IIS) point-to-point links (e.g. T-1, DS-3) are managed now. Traffic from multiple sources over multiple RSVP sessions might be multiplexed on the same VC. This approach has a number of advantages. First, there is typically no signalling latency as VCs would be in existence when the traffic started flowing, so no time is wasted in setting up VCs. Second, the heterogeneity problem (section 4.2.2) in full over ATM has been reduced to a solved problem. Finally, the dynamic QoS problem (section 4.2.7) for ATM has also been reduced to a solved problem.

The aggregation model can be used with point-to-point and point-to-multipoint VCs. The problem with the aggregation model is that the choice of what QoS to use for the VCs may be difficult, without knowledge of the likely reservation types and sizes but is made easier since the VCs can be changed as needed.

4.2.2 Unicast Data VC Management

Unicast data VC management is much simpler than multicast data VC management but there are still some similar issues. If one considers unicast to be a devolved case of multicast, then implementing the multicast solutions will cover unicast. However, some may want to consider unicast-only implementations. In these situations, the choice of using a single flow per VC or aggregation of flows onto a single VC remains but the problem of heterogeneity discussed in the following section is removed.

4.2.3 Multicast Heterogeneity

As mentioned in section 2.1.3.1 and shown in figure 2, multicast heterogeneity occurs when receivers request different qualities of service within a single session. This means that the amount of requested resources differs on a per next hop basis. A related type of heterogeneity occurs due to best-effort receivers. In any IP

multicast group, it is possible that some receivers will request QoS (via RSVP) and some receivers will not. In shared media networks, like Ethernet, receivers that have not requested resources can typically be given identical service to those that have without complications. This is not the case with ATM. In ATM networks, any additional end-points of a VC must be explicitly added. There may be costs associated with adding the best-effort receiver, and there might not be adequate resources. An RSVP over ATM solution will need to support heterogeneous receivers even though ATM does not currently provide such support directly.

RSVP heterogeneity is supported over ATM in the way RSVP reservations are mapped into ATM VCs. There are four alternative approaches this mapping. There are multiple models for supporting RSVP heterogeneity over ATM. Section 4.2.3.1 examines the multiple VCs per RSVP reservation (or full heterogeneity) model where a single reservation can be forwarded onto several VCs each with a different QoS. Section 4.2.3.2 presents a limited heterogeneity model where exactly one QoS VC is used along with a best effort VC. Section 4.2.3.3 examines the VC per RSVP reservation (or homogeneous) model, where each RSVP reservation is mapped to a single ATM VC. Section 4.2.3.4 describes the aggregation model allowing aggregation of multiple RSVP reservations into a single VC.

4.2.3.1 Full Heterogeneity Model

RSVP supports heterogeneous QoS, meaning that different receivers of the same multicast group can request a different QoS. But importantly, some receivers might have no reservation at all and want to receive the traffic on a best effort service basis. The IP model allows receivers to join a multicast group at any time on a best effort basis, and it is important that ATM as part of the Internet continue to provide this service. We define the "full heterogeneity" model as providing a separate VC for each distinct QoS for a multicast session including best effort and one or more qualities of service.

Note that while full heterogeneity gives users exactly what they request, it requires more resources of the network than other possible approaches. The exact amount of bandwidth used for duplicate traffic depends on the network topology and group membership.

4.2.3.2 Limited Heterogeneity Model

We define the "limited heterogeneity" model as the case where the receivers of a multicast session are limited to use either best effort service or a single alternate quality of service. The alternate QoS can be chosen either by higher level protocols or by

dynamic renegotiation of QoS as described below.

In order to support limited heterogeneity, each ATM edge device participating in a session would need at most two VCs. One VC would be a point-to-multipoint best effort service VC and would serve all best effort service IP destinations for this RSVP session.

The other VC would be a point to multipoint VC with QoS and would serve all IP destinations for this RSVP session that have an RSVP reservation established.

As with full heterogeneity, a disadvantage of the limited heterogeneity scheme is that each packet will need to be duplicated at the network layer and one copy sent into each of the 2 VCs. Again, the exact amount of excess traffic will depend on the network topology and group membership. If any of the existing QoS VC endpoints cannot upgrade to the new QoS, then the new reservation fails though the resources exist for the new receiver.

4.2.3.3 Homogeneous and Modified Homogeneous Models

We define the "homogeneous" model as the case where all receivers of a multicast session use a single quality of service VC. Best-effort receivers also use the single RSVP triggered QoS VC. The single VC can be a point-to-point or point-to-multipoint as appropriate. The QoS VC is sized to provide the maximum resources requested by all RSVP next-hops.

This model matches the way the current RSVP specification addresses heterogeneous requests. The current processing rules and traffic control interface describe a model where the largest requested reservation for a specific outgoing interface is used in resource allocation, and traffic is transmitted at the higher rate to all next-hops. This approach would be the simplest method for RSVP over ATM implementations.

While this approach is simple to implement, providing better than best-effort service may actually be the opposite of what the user desires. There may be charges incurred or resources that are wrongfully allocated. There are two specific problems. The first problem is that a user making a small or no reservation would share a QoS VC resources without making (and perhaps paying for) an RSVP reservation. The second problem is that a receiver may not receive any data. This may occur when there is insufficient resources to add a receiver. The rejected user would not be added to the single VC and it would not even receive traffic on a best effort basis.

Not sending data traffic to best-effort receivers because of another receiver's RSVP request is clearly unacceptable. The previously described limited heterogeneous model ensures that data is always sent to both QoS and best-effort receivers, but it does so by requiring replication of data at the sender in all cases. It is possible to extend the homogeneous model to both ensure that data is always sent to best-effort receivers and also to avoid replication in the normal case. This extension is to add special handling for the case where a best-effort receiver cannot be added to the QoS VC. In this case, a best effort VC can be established to any receivers that could not be added to the QoS VC. Only in this special error case would senders be required to replicate data. We define this approach as the "modified homogeneous" model.

4.2.3.4 Aggregation

The last scheme is the multiple RSVP reservations per VC (or aggregation) model. With this model, large VCs could be set up between IP routers and hosts in an ATM network. These VCs could be managed much like IP Integrated Service (IIS) point-to-point links (e.g. T-1, DS-3) are managed now. Traffic from multiple sources over multiple RSVP sessions might be multiplexed on the same VC. This approach has a number of advantages. First, there is typically no signalling latency as VCs would be in existence when the traffic started flowing, so no time is wasted in setting up VCs. Second, the heterogeneity problem in full over ATM has been reduced to a solved problem. Finally, the dynamic QoS problem for ATM has also been reduced to a solved problem. This approach can be used with point-to-point and point-to-multipoint VCs. The problem with the aggregation approach is that the choice of what QoS to use for which of the VCs is difficult, but is made easier if the VCs can be changed as needed.

4.2.4 Multicast End-Point Identification

Implementations must be able to identify ATM end-points participating in an IP multicast group. The ATM end-points will be IP multicast receivers and/or next-hops. Both QoS and best-effort end-points must be identified. RSVP next-hop information will provide QoS end-points, but not best-effort end-points. Another issue is identifying end-points of multicast traffic handled by non-RSVP capable next-hops. In this case a PATH message travels through a non-RSVP egress router on the way to the next hop RSVP node. When the next hop RSVP node sends a RESV message it may arrive at the source over a different route than what the data is using. The source will get the RESV message, but will not know which egress router needs the QoS. For unicast sessions, there is no problem since the ATM end-point will be the IP next-hop router. Unfortunately, multicast routing may

not be able to uniquely identify the IP next-hop router. So it is possible that a multicast end-point can not be identified.

In the most common case, MARS will be used to identify all end-points of a multicast group. In the router to router case, a multicast routing protocol may provide all next-hops for a particular multicast group. In either case, RSVP over ATM implementations must obtain a full list of end-points, both QoS and non-QoS, using the appropriate mechanisms. The full list can be compared against the RSVP identified end-points to determine the list of best-effort receivers. There is no straightforward solution to uniquely identifying end-points of multicast traffic handled by non-RSVP next hops. The preferred solution is to use multicast routing protocols that support unique end-point identification. In cases where such routing protocols are unavailable, all IP routers that will be used to support RSVP over ATM should support RSVP. To ensure proper behavior, implementations should, by default, only establish RSVP-initiated VCs to RSVP capable end-points.

4.2.5 Multicast Data Distribution

Two models are planned for IP multicast data distribution over ATM. In one model, senders establish point-to-multipoint VCs to all ATM attached destinations, and data is then sent over these VCs. This model is often called "multicast mesh" or "VC mesh" mode distribution. In the second model, senders send data over point-to-point VCs to a central point and the central point relays the data onto point-to-multipoint VCs that have been established to all receivers of the IP multicast group. This model is often referred to as "multicast server" mode distribution. RSVP over ATM solutions must ensure that IP multicast data is distributed with appropriate QoS.

In the Classical IP context, multicast server support is provided via MARS [5]. MARS does not currently provide a way to communicate QoS requirements to a MARS multicast server. Therefore, RSVP over ATM implementations must, by default, support "mesh-mode" distribution for RSVP controlled multicast flows. When using multicast servers that do not support QoS requests, a sender must set the service, not global, break bit(s).

4.2.6 Receiver Transitions

When setting up a point-to-multipoint VCs for multicast RSVP sessions, there will be a time when some receivers have been added to a QoS VC and some have not. During such transition times it is possible to start sending data on the newly established VC. The issue is when to start send data on the new VC. If data is sent both on the new VC and the old VC, then data will be delivered with proper

QoS to some receivers and with the old QoS to all receivers. This means the QoS receivers can get duplicate data. If data is sent just on the new QoS VC, the receivers that have not yet been added will lose information. So, the issue comes down to whether to send to both the old and new VCs, or to send to just one of the VCs. In one case duplicate information will be received, in the other some information may not be received.

This issue needs to be considered for three cases:

- When establishing the first QoS VC
- When establishing a VC to support a QoS change
- When adding a new end-point to an already established QoS VC

The first two cases are very similar. In both, it is possible to send data on the partially completed new VC, and the issue of duplicate versus lost information is the same. The last case is when an end-point must be added to an existing QoS VC. In this case the end-point must be both added to the QoS VC and dropped from a best-effort VC. The issue is which to do first. If the add is first requested, then the end-point may get duplicate information. If the drop is requested first, then the end-point may lose information.

In order to ensure predictable behavior and delivery of data to all receivers, data can only be sent on a new VCs once all parties have been added. This will ensure that all data is only delivered once to all receivers. This approach does not quite apply for the last case. In the last case, the add operation should be completed first, then the drop operation. This means that receivers must be prepared to receive some duplicate packets at times of QoS setup.

4.2.7 Dynamic QoS

RSVP provides dynamic quality of service (QoS) in that the resources that are requested may change at any time. There are several common reasons for a change of reservation QoS.

1. An existing receiver can request a new larger (or smaller) QoS.
2. A sender may change its traffic specification (TSpec), which can trigger a change in the reservation requests of the receivers.
3. A new sender can start sending to a multicast group with a larger traffic specification than existing senders, triggering larger reservations.
4. A new receiver can make a reservation that is larger than existing reservations.

If the limited heterogeneity model is being used and the merge node for the larger reservation is an ATM edge device, a new larger reservation must be set up across the ATM network. Since ATM service, as currently defined in UNI 3.x and UNI 4.0, does not allow renegotiating the QoS of a VC, dynamically changing the reservation means creating a new VC with the new QoS, and tearing down an established VC. Tearing down a VC and setting up a new VC in ATM are complex operations that involve a non-trivial amount of processing time, and may have a substantial latency. There are several options for dealing with this mismatch in service. A specific approach will need to be a part of any RSVP over ATM solution.

The default method for supporting changes in RSVP reservations is to attempt to replace an existing VC with a new appropriately sized VC. During setup of the replacement VC, the old VC must be left in place unmodified. The old VC is left unmodified to minimize interruption of QoS data delivery. Once the replacement VC is established, data transmission is shifted to the new VC, and the old VC is then closed. If setup of the replacement VC fails, then the old QoS VC should continue to be used. When the new reservation is greater than the old reservation, the reservation request should be answered with an error. When the new reservation is less than the old reservation, the request should be treated as if the modification was successful. While leaving the larger allocation in place is suboptimal, it maximizes delivery of service to the user. Implementations should retry replacing the too large VC after some appropriate elapsed time.

One additional issue is that only one QoS change can be processed at one time per reservation. If the (RSVP) requested QoS is changed while the first replacement VC is still being setup, then the replacement VC is released and the whole VC replacement process is restarted. To limit the number of changes and to avoid excessive signalling load, implementations may limit the number of changes that will be processed in a given period. One implementation approach would have each ATM edge device configured with a time parameter T (which can change over time) that gives the minimum amount of time the edge device will wait between successive changes of the QoS of a particular VC. Thus if the QoS of a VC is changed at time t , all messages that would change the QoS of that VC that arrive before time $t+T$ would be queued. If several messages changing the QoS of a VC arrive during the interval, redundant messages can be discarded. At time $t+T$, the remaining change(s) of QoS, if any, can be executed. This timer approach would apply more generally to any network structure, and might be worthwhile to incorporate into RSVP.

The sequence of events for a single VC would be

- Wait if timer is active
- Establish VC with new QoS
- Remap data traffic to new VC
- Tear down old VC
- Activate timer

There is an interesting interaction between heterogeneous reservations and dynamic QoS. In the case where a RESV message is received from a new next-hop and the requested resources are larger than any existing reservation, both dynamic QoS and heterogeneity need to be addressed. A key issue is whether to first add the new next-hop or to change to the new QoS. This is a fairly straight forward special case. Since the older, smaller reservation does not support the new next-hop, the dynamic QoS process should be initiated first. Since the new QoS is only needed by the new next-hop, it should be the first end-point of the new VC. This way signalling is minimized when the setup to the new next-hop fails.

4.2.8 Short-Cuts

Short-cuts [4] allow ATM attached routers and hosts to directly establish point-to-point VCs across LIS boundaries, i.e., the VC end-points are on different IP subnets. The ability for short-cuts and RSVP to interoperate has been raised as a general question. An area of concern is the ability to handle asymmetric short-cuts. Specifically how RSVP can handle the case where a downstream short-cut may not have a matching upstream short-cut. In this case, PATH and RESV messages following different paths.

Examination of RSVP shows that the protocol already includes mechanisms that will support short-cuts. The mechanism is the same one used to support RESV messages arriving at the wrong router and the wrong interface. The key aspect of this mechanism is RSVP only processing messages that arrive at the proper interface and RSVP forwarding of messages that arrive on the wrong interface. The proper interface is indicated in the NHOP object of the message. So, existing RSVP mechanisms will support asymmetric short-cuts. The short-cut model of VC establishment still poses several issues when running with RSVP. The major issues are dealing with established best-effort short-cuts, when to establish short-cuts, and QoS only short-cuts. These issues will need to be addressed by RSVP implementations.

The key issue to be addressed by any RSVP over ATM solution is when to establish a short-cut for a QoS data flow. The default behavior is to simply follow best-effort traffic. When a short-cut has been

established for best-effort traffic to a destination or next-hop, that same end-point should be used when setting up RSVP triggered VCs for QoS traffic to the same destination or next-hop. This will happen naturally when PATH messages are forwarded over the best-effort short-cut. Note that in this approach when best-effort short-cuts are never established, RSVP triggered QoS short-cuts will also never be established. More study is expected in this area.

4.2.9 VC Teardown

RSVP can identify from either explicit messages or timeouts when a data VC is no longer needed. Therefore, data VCs set up to support RSVP controlled flows should only be released at the direction of RSVP. VCs must not be timed out due to inactivity by either the VC initiator or the VC receiver. This conflicts with VCs timing out as described in RFC 1755 [11], section 3.4 on VC Teardown. RFC 1755 recommends tearing down a VC that is inactive for a certain length of time. Twenty minutes is recommended. This timeout is typically implemented at both the VC initiator and the VC receiver. Although, section 3.1 of the update to RFC 1755 [11] states that inactivity timers must not be used at the VC receiver.

When this timeout occurs for an RSVP initiated VC, a valid VC with QoS will be torn down unexpectedly. While this behavior is acceptable for best-effort traffic, it is important that RSVP controlled VCs not be torn down. If there is no choice about the VC being torn down, the RSVP daemon must be notified, so a reservation failure message can be sent.

For VCs initiated at the request of RSVP, the configurable inactivity timer mentioned in [11] must be set to "infinite". Setting the inactivity timer value at the VC initiator should not be problematic since the proper value can be relayed internally at the originator. Setting the inactivity timer at the VC receiver is more difficult, and would require some mechanism to signal that an incoming VC was RSVP initiated. To avoid this complexity and to conform to [11] implementations must not use an inactivity timer to clear received connections.

4.3 RSVP Control Management

One last important issue is providing a data path for the RSVP messages themselves. There are two main types of messages in RSVP, PATH and RESV. PATH messages are sent to unicast or multicast addresses, while RESV messages are sent only to unicast addresses. Other RSVP messages are handled similar to either PATH or RESV, although this might be more complicated for RERR messages. So ATM VCs used for RSVP signalling messages need to provide both unicast

and multicast functionality. There are several different approaches for how to assign VCs to use for RSVP signalling messages.

The main approaches are:

- use same VC as data
- single VC per session
- single point-to-multipoint VC multiplexed among sessions
- multiple point-to-point VCs multiplexed among sessions

There are several different issues that affect the choice of how to assign VCs for RSVP signalling. One issue is the number of additional VCs needed for RSVP signalling. Related to this issue is the degree of multiplexing on the RSVP VCs. In general more multiplexing means fewer VCs. An additional issue is the latency in dynamically setting up new RSVP signalling VCs. A final issue is complexity of implementation. The remainder of this section discusses the issues and tradeoffs among these different approaches and suggests guidelines for when to use which alternative.

4.3.1 Mixed data and control traffic

In this scheme RSVP signalling messages are sent on the same VCs as is the data traffic. The main advantage of this scheme is that no additional VCs are needed beyond what is needed for the data traffic. An additional advantage is that there is no ATM signalling latency for PATH messages (which follow the same routing as the data messages). However there can be a major problem when data traffic on a VC is nonconforming. With nonconforming traffic, RSVP signalling messages may be dropped. While RSVP is resilient to a moderate level of dropped messages, excessive drops would lead to repeated tearing down and re-establishing of QoS VCs, a very undesirable behavior for ATM. Due to these problems, this may not be a good choice for providing RSVP signalling messages, even though the number of VCs needed for this scheme is minimized. One variation of this scheme is to use the best effort data path for signalling traffic. In this scheme, there is no issue with nonconforming traffic, but there is an issue with congestion in the ATM network. RSVP provides some resiliency to message loss due to congestion, but RSVP control messages should be offered a preferred class of service. A related variation of this scheme that is hopeful but requires further study is to have a packet scheduling algorithm (before entering the ATM network) that gives priority to the RSVP signalling traffic. This can be difficult to do at the IP layer.

4.3.1.1 Single RSVP VC per RSVP Reservation

In this scheme, there is a parallel RSVP signalling VC for each RSVP reservation. This scheme results in twice the number of VCs, but means that RSVP signalling messages have the advantage of a separate VC. This separate VC means that RSVP signalling messages have their own traffic contract and compliant signalling messages are not subject to dropping due to other noncompliant traffic (such as can happen with the scheme in section 4.3.1). The advantage of this scheme is its simplicity - whenever a data VC is created, a separate RSVP signalling VC is created. The disadvantage of the extra VC is that extra ATM signalling needs to be done. Additionally, this scheme requires twice the minimum number of VCs and also additional latency, but is quite simple.

4.3.1.2 Multiplexed point-to-multipoint RSVP VCs

In this scheme, there is a single point-to-multipoint RSVP signalling VC for each unique ingress router and unique set of egress routers. This scheme allows multiplexing of RSVP signalling traffic that shares the same ingress router and the same egress routers. This can save on the number of VCs, by multiplexing, but there are problems when the destinations of the multiplexed point-to-multipoint VCs are changing. Several alternatives exist in these cases, that have applicability in different situations. First, when the egress routers change, the ingress router can check if it already has a point-to-multipoint RSVP signalling VC for the new list of egress routers. If the RSVP signalling VC already exists, then the RSVP signalling traffic can be switched to this existing VC. If no such VC exists, one approach would be to create a new VC with the new list of egress routers. Other approaches include modifying the existing VC to add an egress router or using a separate new VC for the new egress routers. When a destination drops out of a group, an alternative would be to keep sending to the existing VC even though some traffic is wasted. The number of VCs used in this scheme is a function of traffic patterns across the ATM network, but is always less than the number used with the Single RSVP VC per data VC. In addition, existing best effort data VCs could be used for RSVP signalling. Reusing best effort VCs saves on the number of VCs at the cost of higher probability of RSVP signalling packet loss. One possible place where this scheme will work well is in the core of the network where there is the most opportunity to take advantage of the savings due to multiplexing. The exact savings depend on the patterns of traffic and the topology of the ATM network.

4.3.1.3 Multiplexed point-to-point RSVP VCs

In this scheme, multiple point-to-point RSVP signalling VCs are used for a single point-to-multipoint data VC. This scheme allows multiplexing of RSVP signalling traffic but requires the same traffic to be sent on each of several VCs. This scheme is quite flexible and allows a large amount of multiplexing.

Since point-to-point VCs can set up a reverse channel at the same time as setting up the forward channel, this scheme could save substantially on signalling cost. In addition, signalling traffic could share existing best effort VCs. Sharing existing best effort VCs reduces the total number of VCs needed, but might cause signalling traffic drops if there is congestion in the ATM network. This point-to-point scheme would work well in the core of the network where there is much opportunity for multiplexing. Also in the core of the network, RSVP VCs can stay permanently established either as Permanent Virtual Circuits (PVCs) or as long lived Switched Virtual Circuits (SVCs). The number of VCs in this scheme will depend on traffic patterns, but in the core of a network would be approximately $n(n-1)/2$ where n is the number of IP nodes in the network. In the core of the network, this will typically be small compared to the total number of VCs.

4.3.2 QoS for RSVP VCs

There is an issue of what QoS, if any, to assign to the RSVP signalling VCs. For other RSVP VC schemes, a QoS (possibly best effort) will be needed. What QoS to use partially depends on the expected level of multiplexing that is being done on the VCs, and the expected reliability of best effort VCs. Since RSVP signalling is infrequent (typically every 30 seconds), only a relatively small QoS should be needed. This is important since using a larger QoS risks the VC setup being rejected for lack of resources. Falling back to best effort when a QoS call is rejected is possible, but if the ATM net is congested, there will likely be problems with RSVP packet loss on the best effort VC also. Additional experimentation is needed in this area.

5. Encapsulation

Since RSVP is a signalling protocol used to control flows of IP data packets, encapsulation for both RSVP packets and associated IP data packets must be defined. The methods for transmitting IP packets over ATM (Classical IP over ATM[10], LANE[17], and MPOA[18]) are all based on the encapsulations defined in RFC1483 [19]. RFC1483 specifies two encapsulations, LLC Encapsulation and VC-based multiplexing. The former allows multiple protocols to be encapsulated over the same VC

and the latter requires different VCs for different protocols.

For the purposes of RSVP over ATM, any encapsulation can be used as long as the VCs are managed in accordance to the methods outlined in Section 4. Obviously, running multiple protocol data streams over the same VC with LLC encapsulation can cause the same problems as running multiple flows over the same VC.

While none of the transmission methods directly address the issue of QoS, RFC1755 [11] does suggest some common values for VC setup for best-effort traffic. [14] discusses the relationship of the RFC1755 setup parameters and those needed to support IntServ flows in greater detail.

6. Security Considerations

The same considerations stated in [1] and [11] apply to this document. There are no additional security issues raised in this document.

7. References

- [1] Braden, R., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2209, September 1997.
- [2] Borden, M., Crawley, E., Davie, B., and S. Batsell, "Integration of Realtime Services in an IP-ATM Network Architecture", RFC 1821, August 1995.
- [3] Cole, R., Shur, D., and C. Villamizar, "IP over ATM: A Framework Document", RFC 1932, April 1996.
- [4] Luciani, J., Katz, D., Piscitello, D., Cole, B., and N. Doraswamy, "NBMA Next Hop Resolution Protocol (NHRP)", RFC 2332, April 1998.
- [5] Armitage, G., "Support for Multicast over UNI 3.0/3.1 based ATM Networks", RFC 2022, November 1996.
- [6] Shenker, S., and C. Partridge, "Specification of Guaranteed Quality of Service", RFC 2212, September 1997.
- [7] Wroclawski, J., "Specification of the Controlled-Load Network Element Service", RFC 2211, September 1997.
- [8] ATM Forum. ATM User-Network Interface Specification Version 3.0. Prentice Hall, September 1993.

- [9] ATM Forum. ATM User Network Interface (UNI) Specification Version 3.1. Prentice Hall, June 1995.
- [10] Laubach, M., "Classical IP and ARP over ATM", RFC 2225, April 1998.
- [11] Perez, M., Mankin, A., Hoffman, E., Grossman, G., and A. Malis, "ATM Signalling Support for IP over ATM", RFC 1755, February 1995.
- [12] Herzog, S., "RSVP Extensions for Policy Control", Work in Progress.
- [13] Herzog, S., "Local Policy Modules (LPM): Policy Control for RSVP", Work in Progress.
- [14] Borden, M., and M. Garrett, "Interoperation of Controlled-Load and Guaranteed Service with ATM", RFC 2381, August 1998.
- [15] Berger, L., "RSVP over ATM Implementation Requirements", RFC 2380, August 1998.
- [16] Berger, L., "RSVP over ATM Implementation Guidelines", RFC 2379, August 1998.
- [17] ATM Forum Technical Committee. LAN Emulation over ATM, Version 1.0 Specification, af-lane-0021.000, January 1995.
- [18] ATM Forum Technical Committee. Baseline Text for MPOA, af-95-0824r9, September 1996.
- [19] Heinanen, J., "Multiprotocol Encapsulation over ATM Adaptation Layer 5", RFC 1483, July 1993.
- [20] ATM Forum Technical Committee. LAN Emulation over ATM Version 2 - LUNI Specification, December 1996.
- [21] ATM Forum Technical Committee. Traffic Management Specification v4.0, af-tm-0056.000, April 1996.
- [22] Callon, R., et al., "A Framework for Multiprotocol Label Switching, Work in Progress.
- [23] Rajagopalan, B., Nair, R., Sandick, H., and E. Crawley, "A Framework for QoS-based Routing in the Internet", RFC 2386, August 1998.

- [24] ITU-T. Digital Subscriber Signaling System No. 2-Connection modification: Peak cell rate modification by the connection owner, ITU-T Recommendation Q.2963.1, July 1996.
- [25] ITU-T. Digital Subscriber Signaling System No. 2-Connection characteristics negotiation during call/connection establishment phase, ITU-T Recommendation Q.2962, July 1996.
- [26] ATM Forum Technical Committee. Private Network-Network Interface Specification v1.0 (PNNI), March 1996.

8. Authors' Addresses

Eric S. Crawley
Argon Networks
25 Porter Road
Littleton, Ma 01460

Phone: +1 978 486-0665
EMail: esc@argon.com

Lou Berger
FORE Systems
6905 Rockledge Drive
Suite 800
Bethesda, MD 20817

Phone: +1 301 571-2534
EMail: lberger@fore.com

Steven Berson
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292

Phone: +1 310 822-1511
EMail: berson@isi.edu

Fred Baker
Cisco Systems
519 Lado Drive
Santa Barbara, California 93111

Phone: +1 805 681-0115
EMail: fred@cisco.com

Marty Borden
Bay Networks
125 Nagog Park
Acton, MA 01720

Phone: +1 978 266-1011
EMail: mborden@baynetworks.com

John J. Krawczyk
ArrowPoint Communications
235 Littleton Road
Westford, Massachusetts 01886

Phone: +1 978 692-5875
EMail: jj@arrowpoint.com

9. Full Copyright Statement

Copyright (C) The Internet Society (1998). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

